scientific data

Check for updates

OPEN

DATA DESCRIPTOR

High-Quality Genome Assemblies of Two *Prototheca wickerhamii* Strains

Lihua Fang[1,5], Jian Guo[2,5], Qing Ning[3], Yuhang Luo[4], Jianbo Jian[4] ✉ & Jie Ning[1] ✉

Prototheca wickerhamii **is a non-photosynthetic microalgal species that has been implicated in opportunistic human infections. Understanding its genomic features is crucial for both medical applications and symbiosis research. We generated high-quality genome assemblies for two strains of *Prototheca wickerhamii*, Pw26 and PwS1, using PacBio HiFi reads. The assemblies were evaluated for completeness and accuracy using BUSCO analysis. The assembled genomes for Pw26 and PwS1 were 17.8 MB and 17.4 MB, respectively, with contig N50 values of 1.6 MB. The number of assembled contigs is closely related to the number of chromosomes. The GC content was 63.5% for both genomes. Comparative analysis showed high similarity in genome size and alignment, with Pw26 having slightly more protein-coding genes (46,394) than PwS1 (44,702). Repeat sequences accounted for 6.03% and 4.18% of the genomes in Pw26 and PwS1, respectively. These high-quality genome assemblies provide a valuable resource for comparative genomics and functional exploration of *Prototheca wickerhamii*. The detailed genomic characterization supports further studies on pathogenic mechanisms.**

Background & Summary

Prototheca, a genus of colorless microalgae, is one of the rarely known algae groups causing clinically relevant opportunistic infections in animals and humans[1,2]. Over the past two decades, the incidence of infections caused by these algae, particularly *Prototheca wickerhamii*, has increased significantly. *P. wickerhamii* is the most common species among other *Prototheca* species in opportunistic clinical infections in humans[3,4]. Despite growing attention, there remains a significant gap in our understanding of its biology and pathogenicity[5]. On a global scale, the detection rate of *Prototheca* species infections is significantly lower than the actual prevalence of *P. wickerhamii* infections. The exact mechanisms driving the pathogenesis of *Prototheca* infections are still poorly understood[6].

The first nuclear genome, organelle genomes, and transcriptome of the *P. wickerhamii* type strain ATCC 16529 were publised[4]. The assembled genome size of *P. wickerhamii* (ATCC 16529) is 16.7 Mbp, one of the smallest and most compact among protothecans. A significant portion (23.3%) of genes were annotated with enzymatic activities potentially aiding adaptation to the human host. The genome encodes numerous virulence factors, similar to those in *Candida albicans* and *Trichophyton rubrum*, and approximately 6% of genes matched Pathogen-Host Interaction Database entries[7]. Additionally, genes coding for proteins like ATPase and malate dehydrogenase, previously considered potential virulence factors, were identified. The mitochondrial genome of *Prototheca wickerhamii* has also been previously sequenced at 55,328 base pairs with an A + T content of 74.2%. It encodes essential proteins and RNAs, including cytochrome oxidase subunits, apocytochrome b, NADH dehydrogenase subunits, ATPase subunits, ribosomal RNAs, tRNAs, and ribosomal proteins. Notably, it contains plant-specific mitochondrial genes such as orf25 and rrn5, suggesting a close evolutionary relationship with plant mitochondria[8]. The organellar genomes of true *P. wickerhamii* (type strain ATCC 16529) were also analyzed, and the plastid genome is highly reduced at 48 kb, lacking photosynthesis-related genes but similar to other colorless Prototheca species[9].

[1]Department of Endocrinology, Shenzhen Longhua District Central Hospital, Shenzhen, 518110, China. [2]Department of Laboratory Medicine, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, 200120, China. [3]The Second School of Clinical Medicine, Southern Medical University, Guangzhou, Guangdong Province, 510515, China. [4]Guangdong Provincial Key Laboratory of Marine Biotechnology, Shantou University, Shantou, 515063, China. [5]These authors contributed equally: Lihua Fang, Jian Guo. ✉e-mail: jbjian@126.com; jiening919@gmail.com

Sample	Number of sequences	Sum length	Minimum length	Average length	Maximum length
Pw26	95,701	1,661,276,036	532	17,359	48,048
PwS1	55,962	840,619,978	495	15,021	37,160

Table 1. The sequencing data statistics for the *P. wickerhamii* strain Pw26 and *P. wickerhamii* strain PwS1 genomes.

In a recent study, an atypical mucoid strain of *P. wickerhamii* (S1) was identified and sequenced. Comparative analyses of its transcriptomics, proteomics, and metabolomics with typical strains revealed key differences: S1 exhibited a thinner cell wall due to down-regulation of mannan endo-1,4-β-mannosidase (Maker00002019), reducing macrophage toxicity, and its mucoid appearance was linked to elevated levels of linoleic acid and glycerol[3]. Further research is needed to elucidate the ecology, transmission dynamics, and pathogenesis of *P. wickerhamii* from a One Health perspective.

Given these insights, we are now leveraging third-generation sequencing to explore more genomes of this species, aiming to uncover its broader genomic architecture and evolutionary significance. High-quality genome assemblies are essential for understanding the genetic diversity and evolutionary history of this species[10]. In this study, we aimed to generate high-quality genome assemblies for two strains, Pw26 and PwS1, to provide whole-genome data from two clinical isolates and elucidate their genomic features via comparative genomic analysis[11]. Moreover, functional annotations of protein-coding genes have been conducted, along with investigations into their potential pathogenic roles. By utilizing third-generation sequencing (HiFi) and conducting a comparative analysis of two genomes, we hope to establish a reference information for guiding the clinical treatment of infections caused by this pathogen[12].

Methods

Sample collection, libary construction and sequencing. The two strains of *Prototheca wickerhamii*, namely Pw26 and PwS1, were collected from Shanghai East Hospital by the Protothecosis Science Popularization and Monitor Consortium (PSPMC). These strains were chosen for their relevance in understanding the pathogenic mechanisms and genetic diversity of *Prototheca wickerhamii*. Both *P. wickerhamii* strains were derived from pure, clonal cultures to ensure genomic purity. Strain PwS1 and Pw26 were isolated from two separate clinical cases in Shanghai and Nanning, China, respectively. Each strain was obtained by plating the original clinical sample on Sabouraud Dextrose Agar, incubating at 35 °C for 3 days, and then isolating individual colonies through successive streak-plating to obtain single clones. Purity was confirmed by Gram staining and fluorescence staining, with no bacterial or fungal contamination observed. As reported in the literature for similar *Prototheca* strains, the process of isolating and obtaining the clinical *Prototheca wickerhamii* strains was approved by the Medical Ethics Committee of Shanghai East Hospital, Tongji University (Approval No. 2024YS-274)[13]. Species and strain identification were performed by sequencing the cytochrome b (*cytb*) gene using primers from established references[14] and conducting BLAST analysis against the NCBI database. Therefore, the genomic DNA used for sequencing originated from monoclonal, pure cultures of each strain, confirming that the assembled genomes represent unmixed genotypes.

For the construction of sequencing libraries, high-quality genomic DNA was extracted from agar culture medium of Pw26 and PwS1 using the CTAB method, following a previously published study[5]. The DNA was then sheared into fragments of approximately 20 kb size for sequencing library preparation. Libraries were constructed using the PacBio SMRTbell™ template preparation kit, which involved end-repair, adapter ligation, and exonuclease treatment to remove unligated adapters and small fragments[15]. The sequencing was performed using the PacBio Sequel II platform with the high-fidelity (HiFi) model, which is known for its long-read capability and high accuracy[16]. The raw sequencing data were then processed to remove adapter sequences and low-quality reads, ensuring high-quality data for subsequent assembly and analysis. A total of approximately 1.66 Gb (95,701 reads) and 840.6 Mb (55,962 reads) clean data were generated, respectively (Table 1). The minimum sequence read length was 532 bp for Pw26, while the minimum sequence length was 495 bp for PwS1 (Fig. 1). The average length of CCS reads was 17,359 bp and 15,021 bp for Pw26 and PwS1, respectively (Table 1).

Genome assembly. The long high-quality sequence reads of the two strains were assembled using hifiasm v0.7 with default parameters. The assembled genomes for Pw26 and PwS1 were 17.8 MB and 17.4 MB, respectively, with Contig N50 values of 1.6 Mb (Table 2). The GC content was 63.5% for both genomes(Fig. 4). The assembled genome size for Pw26 was 17,790,715 bp, with a maximal contig length of 2,456,030 bp and a Contig N50 value of 1,637,018 bp. The total number of contigs was 17. For PwS1, the assembled genome size was 17,442,376 bp, with a maximal contig length of 2,486,542 bp and a N50 value of 1,639,725 bp. The total number of contigs was 14.

Repeat annotation and gene annotation. Repeat sequences play a crucial role in shaping the genomic architecture of *Prototheca wickerhamii*. In this study, we annotated the repeat sequences in the genomes of two strains, Pw26 and PwS1. The repeat sequences were identified with TRF (4.9) with default parameters and together with Tandem repeats[17]. Then, the homologous sequences of the Pw26 and PwS1 genomes were annotated using RepeatProteinMask (v 4.0.7)[18] and RepeatMasker (open-4.0.9)[19] combing with the Repbase library (http://www.girinst.org/repbase)[20]. The repetitive sequence database of Pw26 and PwS1 were generated using RepeatModeler open-1.0.1140 and LTR_FINDER_parallel 1.0.741, then the ab initio method was used for repeat identification with RepeatMasker (open-4.0.9)[19]. The repeat sequences accounted for 6.03% of the Pw26 genome

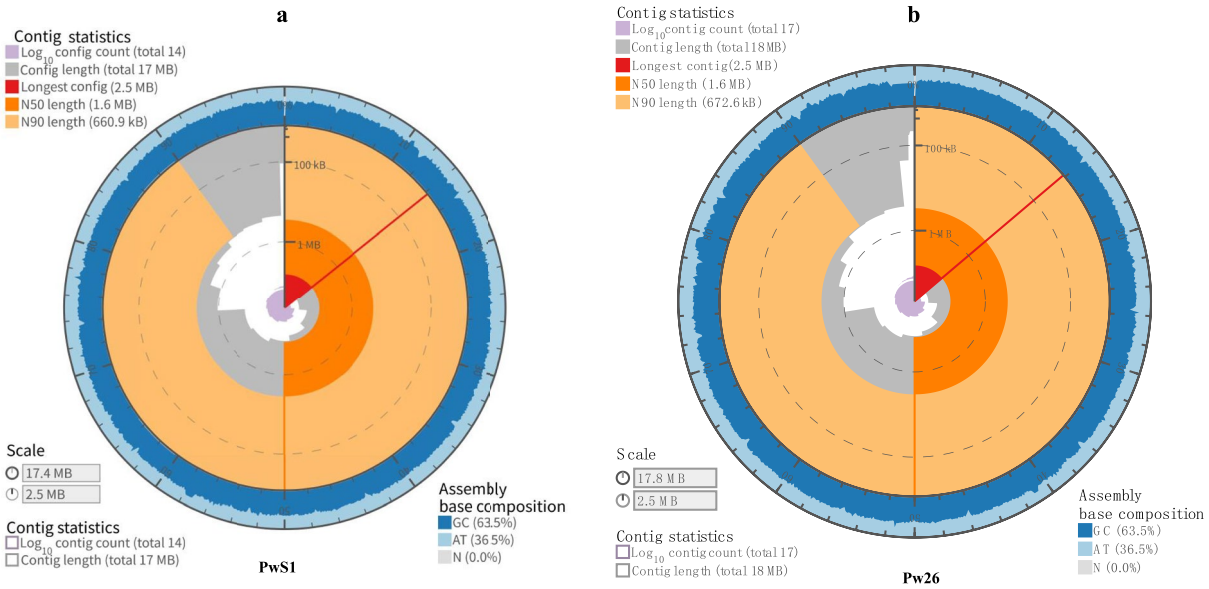

Fig. 1 Snail plot of genome assembly statistics for *Prototheca wickerhamii* Strains Pw26 and PwS1.

Sample	Pw26	PwS1	ATCC 16529*
Total_length	17,790,715	17,442,376	16,700,629
Maximal length	2,456,030	2,486,542	2,447,261
N50	1,637,018	1,639,725	1,578,614
N50 Number	5	5	5
N90	672,571	660,865	616,462
N90 Number	11	11	11
Number of Contig/Scaffold	17	14	19
GC_rate	63.50%	63.50%	62.70%
Gene number	6,443	6,483	6,081
Repeat (%)	6.03	4.18	2.25

Table 2. The genome assembly and annotation statistics for *P. wickerhamii* strain Pw26, strain PwS1 and strain ATCC 16529. *ATTC 16529 was assembled at the scaffold level.

Strains	Complete	Complete and single-copy	Complete and duplicated	Fragmented	Missing
Pw26	87.00%	83.70%	3.30%	0.50%	12.50%
PwS1	88.40%	87.30%	1.10%	0.30%	11.30%

Table 3. The BUSCO statistics of gene set in *P. wickerhamii* strain Pw26 and strain PwS1.

and 4.18% of the PwS1 genome. The majority of these repeats were long terminal repeats (LTRs), which are common in plant or algae genome which contribute to genomic diversity and evolution. The most abundant repeat types were Retro/LTR/Gypsy, with 191,363 bp (1.08%) in Pw26 and 189,820 bp (1.09%) in PwS1. Other notable repeat types included Retro/LTR/Copia, Retro/SINE, and Retro/LINE, which were present in both strains but with slight differences in their proportions (Table 4). The Pw26 strain exhibited a higher overall repeat content than PwS1, which may reflect differences between the two strains and the selective pressures exerted by their respective environments.

After repeat annotation, three different strategies, namely *ab initio* prediction annotation, homology-based annotation and transcriptome-based annotation were used for gene prediction. In the *ab initio* method, the *P. wickerhamii* Pw26 and PwS1 genome sequences, after masked with repeated elements, the coding regions of genes were identified using AUGUSTUS v3.2.3[21]. In homology-based method, a total of 8 published homology protein sequences from three strains of *Prototheca wickerhamii* ATTC16529, S1 and S931, *Chlorella variabilis*,*Chlorella desiccate*, *Coccomyxa subellipsoidea, Chlorella sorokiniana*, *Auxenochlorella protothecoides* were integrated with the MAKER2 pipeline[22]. For transcriptome-based prediction, the published transcriptome of *Prototheca wickerhamii* from previously study[5]. The RNA reads were aligned onto the genomes of two strains of Pw26 and PwS1 using hista2.2.1[23], then transcript was assembled using stringtie2.1.6[24]. The gene sets of two

Type/Sample	Pw26		PwS1	
	Length(bp)	% of genome	Length(bp)	% of genome
Retro/LTR/Copia	8,927	0.05	7,603	0.04
Retro/LTR/Gypsy	191,363	1.08	189,820	1.09
Retro/LTR/Other	8,426	0.05	7,671	0.04
Retro/SINE	13,345	0.08	11,109	0.06
Retro/LINE	20,745	0.12	17,308	0.10
Retro/Other	0	0.00	0	0.00
DNA/EnSpm	8,525	0.05	10,127	0.06
DNA/Harbinger	4,298	0.02	4,006	0.02
DNA/hAT	29,733	0.17	25,679	0.15
DNA/Helitron	3,574	0.02	3,439	0.02
DNA/Mariner	0	0.00	48	0.00
DNA/MuDR	3,243	0.02	3,944	0.02
DNA/P	733	0.00	733	0.00
DNA/Other	23,105	0.13	12,733	0.07
Other	255,507	1.44	17,877	0.10
Unknown	567,452	3.19	459,364	2.63
Total	1,072,400	6.03	729,621	4.18

Table 4. The repetitive sequence statistics in the *P. wickerhamii* strain Pw26 and strain PwS1.

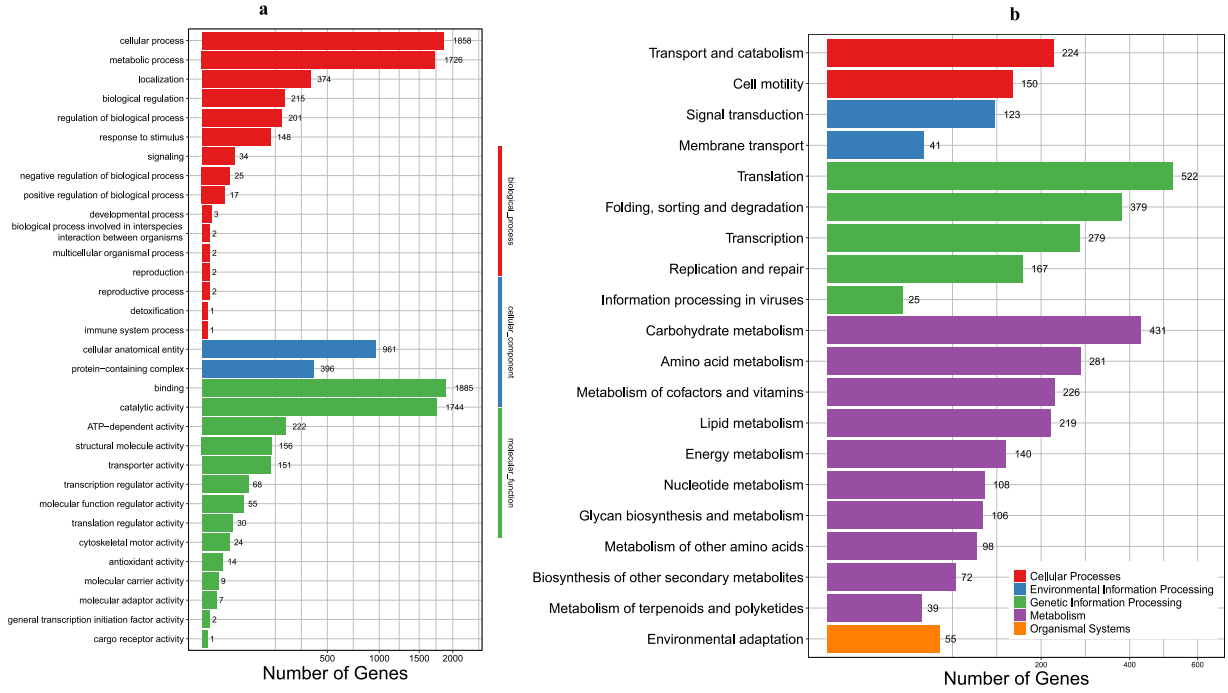

Fig. 2 GO and KEGG Functional Annotation of Protein-Coding Genes in Pw26.

strains were obtained using Maker2 (2.31.10)[22]. Finally, a total of 6443 and 6483 protein-coding genes were obtained respectively for *P. wickerhamii* strain Pw26 and PwS1 genomes (Table 1).

Despite the slight difference in gene number, both strains exhibited high completeness in their gene sets, as assessed by BUSCO analysis (Table 3). This comprehensive annotation provides a valuable resource for further comparative genomic studies and functional exploration of *Prototheca wickerhamii*.

Functional annotation. To elucidate the functional of the two *P. wickerhamii* strains, we conducted functional annotation and pathway analysis using Gene Ontology (GO)[25] and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases[26]. This approach involved detailed annotation of protein-coding genes to identify their functions and the metabolic pathways[27]. The GO analysis[28] revealed a comprehensive distribution of gene functions across various biological processes, molecular functions, and cellular components. For Pw26, the GO annotation identified a diverse range of gene functions, with significant representation in metabolic processes, cellular component organization, and biological regulation. Similarly, the PwS1 strain exhibited a comparable

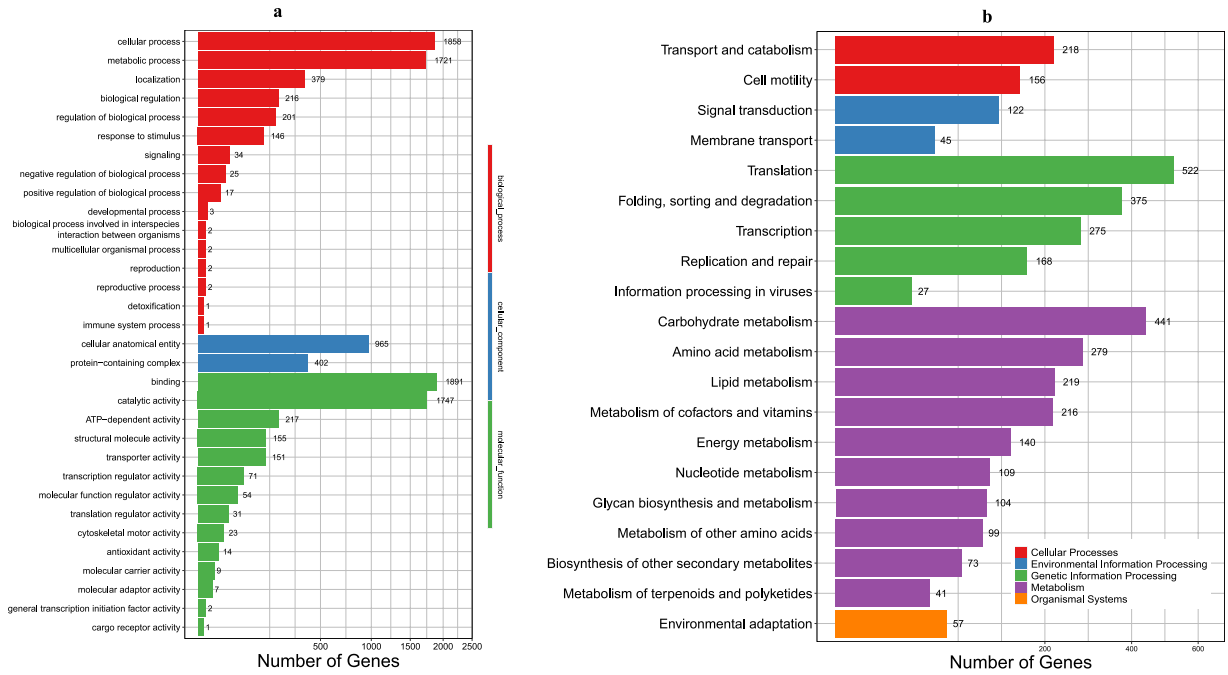

Fig. 3 GO and KEGG Functional Annotation of Protein-Coding Genes in PwS1.

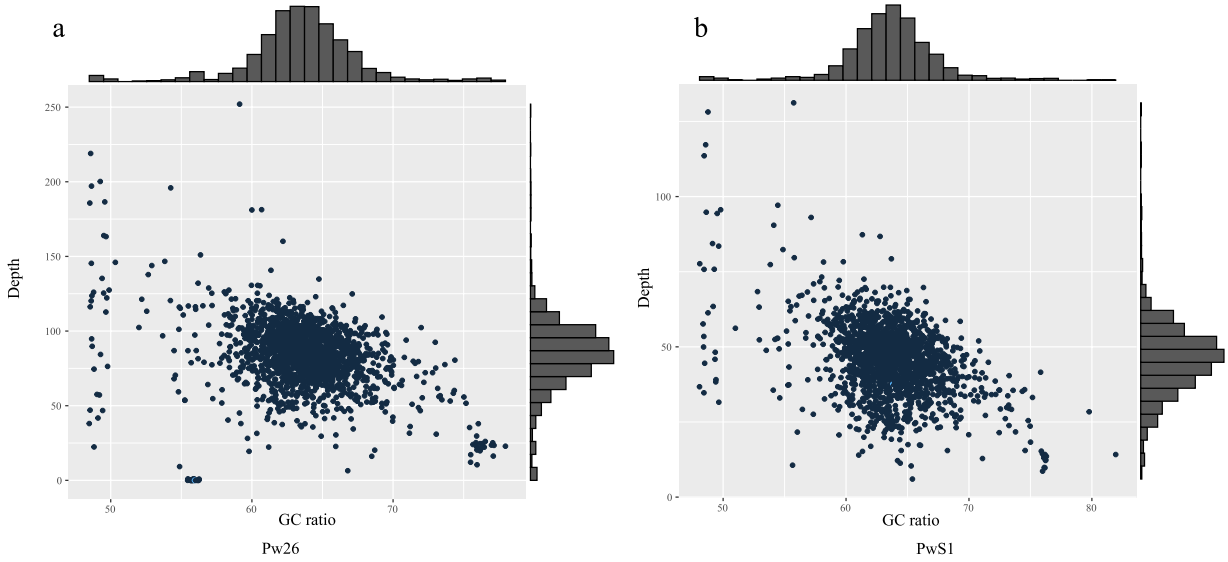

Fig. 4 GC content and depth distribution for HiFi reads in *P. wickerhamii* Strains Pw26 and PwS1.

distribution of gene functions, highlighting key roles in cellular metabolism and regulation(Fig. 2). The KEGG pathway analysis[29] provided insights into the metabolic pathways and biological processes encoded by the genomes(Fig. 3). For Pw26, the KEGG annotation highlighted pathways involved in carbohydrate metabolism, amino acid biosynthesis, and energy production. In PwS1, the KEGG analysis also emphasized pathways related to lipid metabolism and signal transduction, indicating potential differences in metabolic capabilities between the two strains

Data Records

The sequencing datasets, including HiFi long reads for *P. wickerhamii* strains Pw26 (SRR35231276)[30] and PwS1 (SRR35231275)[31], have been deposited in the Sequence Read Archive (SRA) under project number PRJNA1314384. The final, assembled genome sequences have been deposited in NCBI GenBank under accession numbers JBTKVZ000000000 (Pw26)[32] and JBTKVY000000000 (PwS1)[33]. The genome sequences and annotations for these strains are available on figshare[34] at https://doi.org/10.6084/m9.figshare.30030796.v1.

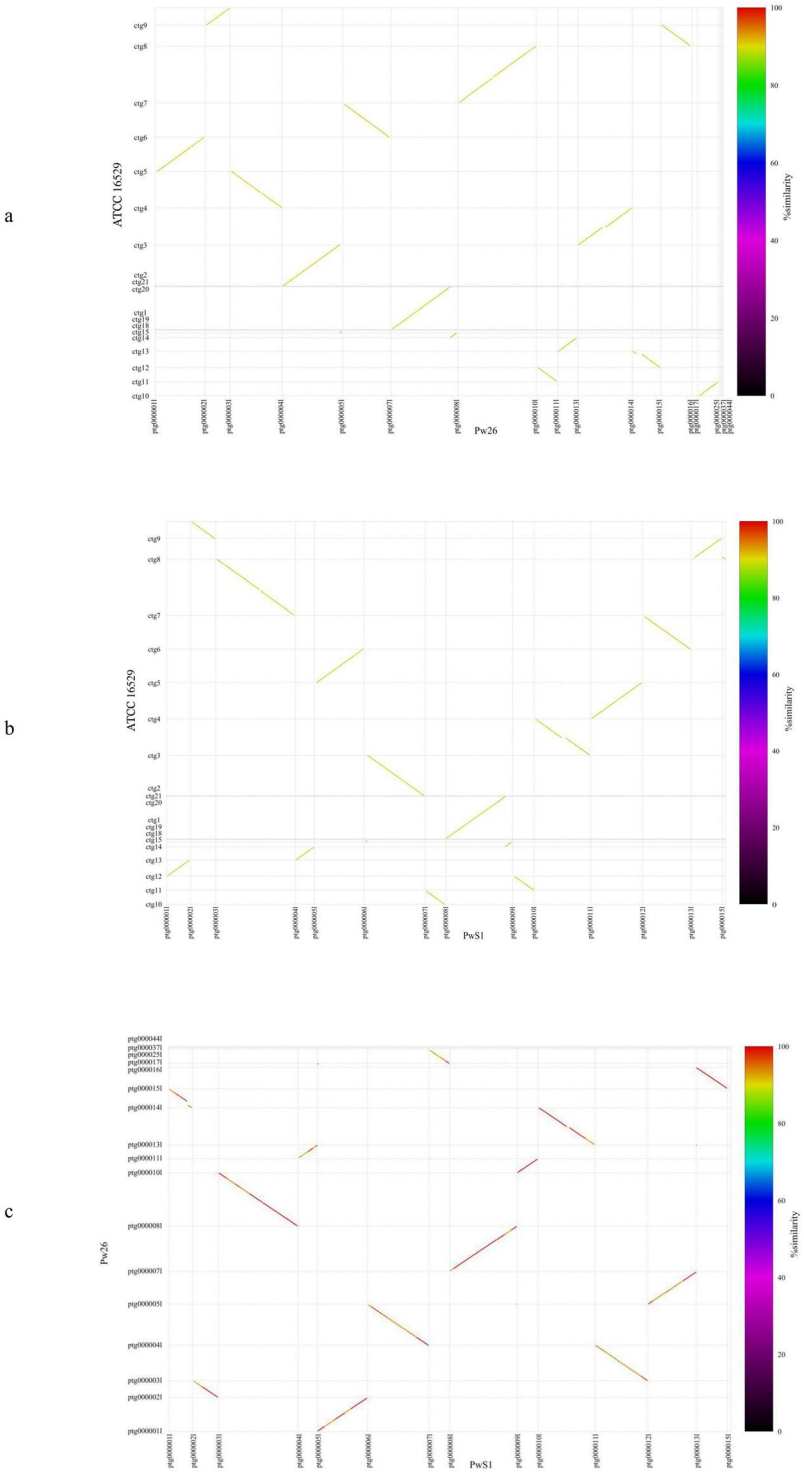

Fig. 5 The genome synteny plot of newly assembled *P. wickerhamii* Strains Pw26, PwS1 and ATCC 16529.

Technical Validation

The quality assessment of the *P. wickerhamii* Pw26 and PwS1 strain genomes was evaluated by mapping the HiFi reads to the assembled genomes using Minimap2[35], resulting in mapping rates of 93.8% and 95.68% for Pw26 and PwS1, respectively, with nearly 100% coverage. The GC content and depth distribution of mapped reads[36] showed a GC content of 63.5% with no significant contamination, indicating high-quality genome assemblies (Fig. 5a).

The two Pw26 and PwS1 genomes were compared using MUMmer4[37] with the nucmer model (c 2000 -l 400). The MUMmer analysis of the *P. wickerhamii* genomes revealed high similarity, suggesting that the genome

assembly was accurate (Fig. 5b). The comparison between the two genomes showed nearly one-to-one contig alignment with high similarity, further supporting the correctness of the genome assembly (Fig. 5c).

Based on the "chlorophyta_odb10" database, the BUSCO analysis[38] revealed high completeness for both genomes, with more than 86% complete BUSCOs for both Pw26 and PwS1. This indicates that the assemblies are highly representative of the entire genomes. For Pw26, 87.00% of the BUSCO genes were complete, with 83.70% being single-copy and 3.30% being duplicated. The fragmented and missing BUSCO genes were 0.50% and 12.50%, respectively[39]. For PwS1, 88.40% of the BUSCO genes were complete, with 87.30% being single-copy and 1.10% being duplicated. The fragmented and missing BUSCO genes were 0.30% and 11.30%, respectively. These results suggest that both Pw26 and PwS1 have complete and well-preserved gene sets, which could be important for understanding their biological functions and potential pathogenic mechanisms.

Data availability

All data related to the genome of *P. wickerhamii* strain PwS1 and Pw26 are available through the following databases or links. Sequence Read Archive (SRA) data was uploaded in NCBI with project ID of PRJNA1314384. The HiFi long reads sequencing data for *P. wickerhamii* Pw26 and PwS1 have been deposited in the SRA, with accession numbers SRR35231276 and SRR35231275, respectively. The whole-genome shotgun projects for strains Pw26 and PwS1 have been deposited in GenBank under accessions JBTKVZ000000000[32] and JBTKVY000000000[33], correspondingly. Additionally, SRA data link is https://identifiers.org/ncbi/insdc.sra:SRP616729. The Figshare link: https://doi.org/10.6084/m9.figshare.30030796.v1.

Code availability

This study did not involve the development of any specific code. The data analyses were conducted in accordance with the protocols outlined in the Methods section.

Received: 8 September 2025; Accepted: 12 February 2026;
Published online: 09 March 2026

References

1. Masuda, M. *et al*. Protothecosis in Dogs and Cats-New Research Directions. *Mycopathologia.* **186**(1), 143–152 (2021).
2. Kano, R. Emergence of Fungal-Like Organisms: Prototheca. *Mycopathologia.* **185**(5), 747–754 (2020).
3. Guo, J. *et al*. Integration of transcriptomics, proteomics, and metabolomics data for the detection of the human pathogenic Prototheca wickerhamii from a One Health perspective. *Frontiers in cellular and infection microbiology.* **13**, 1152198 (2023).
4. Bakuła, Z. *et al*. A first insight into the genome of Prototheca wickerhamii, a major causative agent of human protothecosis. *BMC genomics.* **22**(1), 168 (2021).
5. Guo, J. *et al*. Genome Sequences of Two Strains of Prototheca wickerhamii Provide Insight Into the Protothecosis Evolution. *Frontiers in cellular and infection microbiology.* **12**, 797017 (2022).
6. Lass-Flörl, C. & Mayr, A. Human protothecosis. *Clinical microbiology reviews.* **20**(2), 230–242 (2007).
7. Urban, M. *et al*. PHI-base: the pathogen-host interactions database. *Nucleic acids research.* **48**(D1), D613–d620 (2020).
8. Wolff, G., Plante, I., Lang, B. F., Kück, U. & Burger, G. Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii. Gene content and genome organization. *Journal of molecular biology.* **237**(1), 75–86 (1994).
9. Bakuła, Z. *et al*. Sequencing and Analysis of the Complete Organellar Genomes of Prototheca wickerhamii. *Frontiers in plant science.* **11**, 1296 (2020).
10. Zhang, Q. Q., Zhu, L. P., Weng, X. H., Li, L. & Wang, J. J. Meningitis due to Prototheca wickerhamii: rare case in China. *Medical mycology.* **45**(1), 85–88 (2007).
11. Li, J., Huang, Z. & Zhang, R. Unmasking Prototheca wickerhamii: A rare case of cutaneous infection and its implications for clinical practice. *The Brazilian journal of infectious diseases: an official publication of the Brazilian Society of Infectious Diseases.* **29**(3), 104525 (2025).
12. Etchecopaz A. N., Del Vecchio L., Álvarez C., Mesplet M. & Cuestas M. L. Cytological and microbiological analysis of a Prototheca wickerhamii infection in a cat with cutaneous lesions successfully treated with intralesional amphotericin B. *The Journal of small animal practice* (2025).
13. Guo J. *et al*. Two high-quality genomes of Prototheca bovis strain SH08 and Prototheca ciferrii strain SH13. *Scientific data*. (2025).
14. Jagielski, T. *et al*. Occurrence of Prototheca Microalgae in Aquatic Ecosystems with a Description of Three New Species, Prototheca fontanea, Prototheca lentecrescens, and Prototheca vistulensis. *Applied and environmental microbiology.* **88**(22), e0109222 (2022).
15. Mareso, C. *et al*. Optimization of long-range PCR protocol to prepare filaggrin exon 3 libraries for PacBio long-read sequencing. *Molecular biology reports.* **50**(4), 3119–3127 (2023).
16. Ortigas-Vasquez, A. *et al*. High-fidelity long-read sequencing of an avian herpesvirus reveals extensive intrapopulation diversity in tandem repeat regions. *PLoS pathogens.* **21**(8), e1013435 (2025).
17. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**(2), 573–580 (1999).
18. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*. **6**, 11 (2015).
19. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics.* **21**(Suppl 1), i351–358 (2005).
20. Jurka, J. *et al*. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research.* **110**(1-4), 462–467 (2005).
21. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics.* **7**, 62 (2006).
22. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics.* **12**(1), 491 (2011).
23. Yang, Z. *et al*. Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nature Plants.* **5**(9), 991–1001 (2019).
24. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**(1), 278 (2019).
25. Zhu, M., Wang, X. & Li, X. Genome-wide identification and expression analysis of glutamate receptor-like genes in three Dendrobium species. *Biochimica et biophysica acta General subjects.* **1869**(6), 130789 (2025).
26. Zuo, W. & Wang, Z. Identification of ulcerative colitis diagnostic markers from differentially expressed genes shared with Hirschsprung disease. *Scientific reports.* **15**(1), 11274 (2025).

27. Borza, T., Popescu, C. E. & Lee, R. W. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga Prototheca wickerhamii. *Eukaryotic cell.* **4**(2), 253–261 (2005).
28. Zou, C. *et al.* Genome and transcriptome wide association study identify candidate genes regulating folate levels in maize. *Frontiers in plant science.* **16**, 1606220 (2025).
29. Zhu, J. *et al.* Genome-wide association study and transcriptomic analysis reveal the crucial role of sting1 in resistance to visceral white-nodules disease in Larimichthys polyactis. *Frontiers in immunology.* **16**, 1562307 (2025).
30. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR35231276 (2026).
31. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR35231275 (2026).
32. Fang, L. *et al.* High-Quality Genome Assemblies of Two Prototheca wickerhamii Strains, https://www.ncbi.nlm.nih.gov/nuccore/JBTKVZ000000000 (2026).
33. Fang, L. *et al.* High-Quality Genome Assemblies of Two Prototheca wickerhamii Strains, https://www.ncbi.nlm.nih.gov/nuccore/JBTKVY000000000 (2026).
34. Fang, L. *et al.* High-Quality Genome Assemblies of Two Prototheca wickerhamii Strains. *figshare* https://doi.org/10.6084/m9.figshare.30030796.v1 (2026).
35. Zhou, Q. *et al.* Telomere-to-telomere gapless genome assembly of the giant grouper (Epinephelus lanceolatus). *Scientific data.* **11**(1), 1342 (2024).
36. Zuo, W. *et al.* Whole genome sequencing of a multidrug-resistant Bacillus thuringiensis HM-311 obtained from the Radiation and Heavy metal-polluted soil. *Journal of global antimicrobial resistance.* **21**, 275–277 (2020).
37. Qian, W. *et al.* Identification of novel single nucleotide variants in the drug resistance mechanism of Mycobacterium tuberculosis isolates by whole-genome analysis. *BMC genomics.* **25**(1), 478 (2024).
38. Zhou, X., Wang, E., Xu, X. & Zhang, B. Chromosome-level genome assembly of Phytoseiulus persimilis Athias-Henriot. *Scientific data.* **12**(1), 293 (2025).
39. Sodmann, A. *et al.* Human dorsal root ganglia are either preserved or completely lost after deafferentation by brachial plexus injury. *British journal of anaesthesia.* **133**(6), 1250–1262 (2024).

Acknowledgements

This work was financially supported by the STU Scientific Research Initiation Grant (NTF25030T), Sanming Project of Shenzhen Longhua Distrcit Central Hospital and Municipal Financial Subsidy of Shenzhen Longhua District Key Medical Discipline Construction.

Author contributions

J. Jian conceived the study. J. Guo collected the samples, conducted experiments, L. Fang, J. Jian and Y. Luo performed bioinformatics analysis. L. Fang, Q. Ning, J. Jian wrote the manuscript. J. Guo and J. Ning provided suggestion and revised the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.J. or J.N.